

ORIGINAL ARTICLE

Medicine Science

A supporting method to detect manipulated zones in digitally edited audio files**Ege Niyazi Gural, Melih Pazarci***Istanbul University, Institute of Forensic Sciences, Cerrahpasa, Fatih, Istanbul, Turkey
Istanbul Technical University, Faculty of Electrical, Istanbul, Turkey*Received 2 October 2017; Accepted 18 December 2017
Available online 18.12.2017 with doi: 10.5455/medscience.2017.06.8699**Abstract**

The courts request the forensic audio laboratory to determine whether the digital audio files are original or modified, where the altered files have been modified, and which parts can be used as evidence. It is important to identify manipulation zones in audio files that contain speech as well as to determine the authenticity of the audio file. For this reason, when we examine the theory and rules of digitizing or digital recording, it has been found that inevitable manipulation traces must be found at the manipulation points, but these traces can be removed or reduced by certain software although some manipulations can be re-revealed. In this paper, a supporting method that uncovers the wide band oscillations signals known as the Gibbs effect caused by tape tampering but may have been obscured by software means is given, and it can distinguish such manipulations from other deceptive but benign transient signals. The procedure is applied on digital files in a computer. Lower frequencies are de-emphasized compared to higher frequencies in a special way by software so that manipulation points are easier to detect. Examples are presented. The aim is to reduce forensic analyst investigation time and fatigue. It has been used as a supporting method in our institution since 1998.

Keywords: Forensic audio, digital editing, authentication, manipulated audio files, audio signal analysis**Introduction**

According to Fourier's theorem, a signal can be represented as an infinite sum of sinusoidal signals. If the signal is periodic, the Fourier series can express the periodic signal. We can separate the sound waves into its harmonic components with the Fourier series by considering a finite segment of this sound signal as periodic. Thus a sound wave consists of an infinite number of sinusoidal waves at exactly the multiples of the fundamental frequency. This fundamental frequency depends on the reciprocal of the length of the time segment, hence the longer the segment of the time wave, the higher the spectral resolution, i.e., the smaller the fundamental frequency. Here, the infinite number is a requirement for smooth signal continuity. If any interruption in the signal occurs, discontinuity and reaction will occur as a result. When we cut the sound suddenly, i.e. in a manner not conforming to the underlying audio generation process, such as cut, delete, or paste, a discontinuity will occur. Gibbs effect will occur at this point. In visual examinations, the Fast Fourier Transform (FFT) algorithm will show Gibbs oscillations similarly, as the Discrete Fourier Transform is applied to the discrete signals [1-3].

Acoustic sound is in continuous and periodic forms. It can be represented using Fourier techniques. However, the sound is to be digitized when converting for analysis in a computer since computers can only work with discrete and finite values. For this purpose, sound waves are sampled and digitized to discrete sample values. After working on these discrete samples, at the end, they are to be reconstructed to create analog continuous sound signals that are very close to the acoustic sound. Digital signals are made "visible" in the frequency domain by using the Fast Fourier Transform algorithm. Audio signal manipulations, such as a cutting operation made with the help of software, generate Gibbs oscillations signals due to the discontinuity in the cutting zone. Gibbs oscillations signals have higher energy and a wider frequency band than the signals at the cutting point [1].

Without a cover-up, the traces of cutting and combining can easily be seen and heard except for cutting and pasting in the exact zero-crossing areas and with slope compatibility. Software interpolates and smoothens the remaining detectable marks at the point of interruption, making detection process harder, but with both spectral visual and time domain aural losses. The higher the energy of the signals at the cutting and combining locations, the higher the spectral visibility and audibility of Gibbs oscillations become. Occasionally, Gibbs oscillations that are created by cutting or combining similar transient signals can be camouflaged by in between sound signals and can stand undetected.

*Corresponding Author: Ege Niyazi Gural, Istanbul University, Institute of Forensic Sciences, Cerrahpasa, Fatih, Istanbul, Turkey

E-mail: egegural@istanbul.edu.tr

The method focuses on decreasing the energy in the frequency range where the sound signals are naturally densely located, and increasing the energy of inaudible frequency ranges [1]. Traces of Gibbs oscillations will be observed as thin, sharp lines on the spectrogram. This response may appear as an increase in the harmonics at the interrupted area and an increase in their amplitude, similar to usual transient signals, but distinctly sharper and spectrally not much spread. Later, these detected suspicious locations will be investigated by an analyst with available auditory and visual examination techniques to distinguish them from other transients.

In manipulation software, these Gibbs oscillations are smoothed out as much as possible, even if not completely cleaned. The spectral lines of the remaining oscillation signal traces can be observed in the spectrogram. In some cases it can be heard as a metallic sound even if it cannot be observed in the spectrogram.

In this paper, we propose a method called the Enhanced Audio files Examination Method (EAEM) which enhances the visibility of such Gibbs-effect contaminated manipulation points to increase their visibility to the analyst in order to reduce analysis time and operator fatigue.

Material and Methods

More than digitized 300 audio files including real cases have been tested since 1998 for the development of the method. In this paper, the first example is a compact cassette whose audio content was digitally edited on a computer before being recorded to the cassette. The audio content on this cassette was digitized again for examination. Since some digital editing traces (Gibbs oscillations) may disappear in the electronic background white noise ("hiss" sound), conversion of the digitally edited sound files to compact cassettes was used to make the process more difficult [4].

The first example is a Sony Walkman recording. To assess the advantages of the supportive method, a Sony Walkman recording that has background sound, usual device noises, and usual conversations is examined for visual and auditory review.

Another example is a sound file that contains a manipulated digital recording of a TV commercial which has also undergone noise reduction that reduces manipulation traces; the manipulations and recording are done in a professional studio using high quality professional equipment.

For the examination, we have used computers with sound cards, and the software to examine the signals was Adobe Audition [5].

Enhanced Audio files Examination Method (EAEM)

The method is applied to audio files that are edited by deleting, cutting and pasting.

The examinations are done visually on sound files digitized at a sample rate of 192 kHz (oversampled), then by applying EAEM and finally analyzed in the time domain (waveform) and in the spectral domain (spectrogram). The oversampling is primarily for smoother waveform and spectrogram results at the end. The oversampled signal is first expanded 3:1 (<10dB). This step helps in improving the manipulation- traces-to-original (i.e., when the original source recording is done) noise ratio, hence increases

visibility of traces. The following step is the frequency emphasis.

In the high frequency region, the spectral density of human speech naturally decays. This high frequency range starts somewhere in the 4-10 kHz frequency depending on the voice content. In practice, we mark the frequency where the audio spectral density drops relatively sharply, and suppress the low range by up to 45dB. At frequencies below about 10 kHz, the sharp-spin filter reduces the intensity by 45 dB, after which the entire audio frequency band is normalized. These constitute the first step of the EAEM. In the resulting audio, speech can still be heard, but the suspicious marks are emphasized. By this method, candidate manipulation zones can be determined easier excluding many misleading points. Fused traces, which are not seen in EAEM unapplied audio files, become more pronounced as a result of the EAEM application.

The candidate manipulation locations in both waveform and FFT spectrogram form are visually detected in the EAEM file and the manipulation traces may be observed in one or both domains. In some cases, traces are observed in the spectrogram, whereas in other cases they are observed through waveforms. Occasionally, traces are observed in both the spectrogram and waveforms. Both domains must be considered and examined. The segments where suspicious potential zones are found are also subjected to auditory examination.

In an additional step, the detected candidate manipulation zones are investigated after applying the hard-limiting algorithm (maximum amplitude limit -1 dB, boost input by 50 dB, look ahead time 5 ms, release time 40 ms). The resulting hard limited signals are investigated either spectrographically (this is the second step of the EAEM) or by critical listening (this is the third step of the EAEM).

The first step is generally sufficient in the determination of the existence of manipulations and gives an estimate of the type of the work done on the audio. The second step is necessary for the determination of the well hidden traces such as in very silent time slots but the examination will be slower. EAEM does not make the classic listening, waveform and spectrographic analyses unnecessary, but is a supplementary time saving aid to the investigator. Some forensic audio material may be long in the hours range. In typical use, the investigator will be listening to the original audio while observing the EAEM output. This is best done working in 2-channel mode, where one channel is the original, and the second channel is EAEM applied.

Results

Although the developed method has proven successful on audio recordings with a background (noise or else), in studio recordings with no background, the method resulted with an increased number of false positives. The EAEM detected marks found are considered as candidate manipulation zones to be investigated for acoustical integrity by standard methods. Figure 1 illustrates the Gibbs effect in the spectrogram.

Fig.1 is a spectrogram image of a computer generated stereo audio file that has been edited. In the left channel, a 1 kHz and in the right channel a 10 kHz sinusoidal signal were produced by a computer and subsequently edited in the form of segment deleting. Point (A)

is a random deleted point without a smoothing application. Point (B) is a random deleted point with smoothing application. Point (C) has deleted signals at both left and right channels, however the left channel deletion in the waveform has observed the zero-crossing point with sinus slope compatibility, and smoothing was applied to both channels. In such a clean audio example, the different levels of manipulations have spectrographic traces visible as shown.

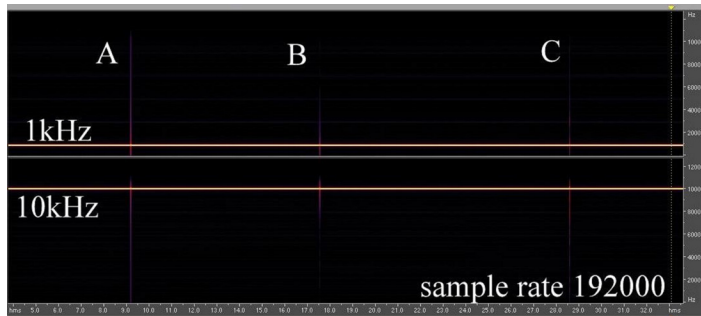


Figure 1. Spectrogram image of a computer generated edited stereo audio file containing 1kHz and 10kHz sinusoidal signals.

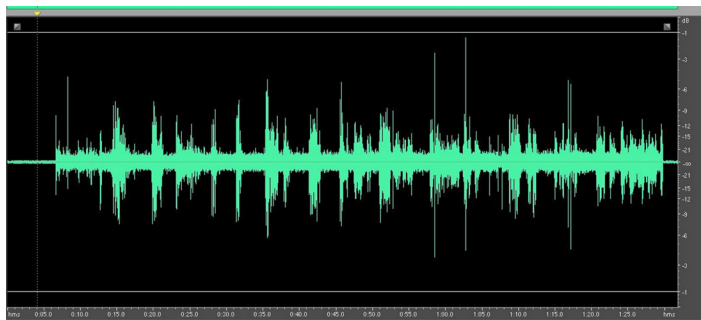


Figure 2. This is a time domain waveform view of the Sony Walkman cassette tape recording containing edited natural environment speech

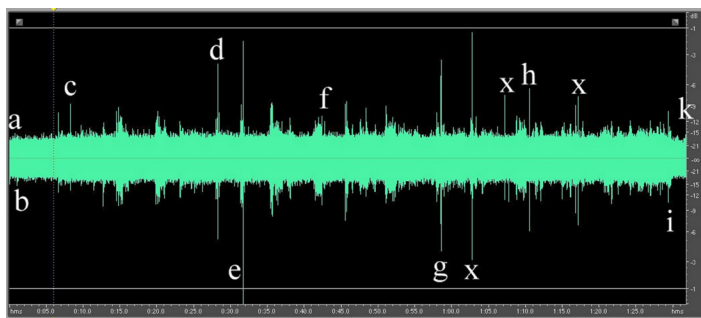


Figure 3. EAEM 1st step resulting waveform of audio file shown in Figure 2. Zones a, b, c, d, e, f, g, h, i, k have been edited

Fig. 3 shows the EAEM 1st step resulting waveform of audio file shown in Fig. 2. Zones a, b, c, d, e, f, g, h, i, k have been edited. The zones that are marked with X are the strong acoustic action sound points such as impacts. a and b are the places where deleting operations were performed on the quietest place of the tape where only “hiss” sound exists. In spite of the fact that these zones are not suitable to zero-crossing, the edits cannot be determined in spectrogram or waveform since they are fused and immersed in the “hiss” noise. d and e are the zones of in-word cut/erase and recombination (e.g., “Kapı aç-ık pencere aç-ık”, where the italic segment has been erased and remains combined). The other edit points do not have any particular specialty.

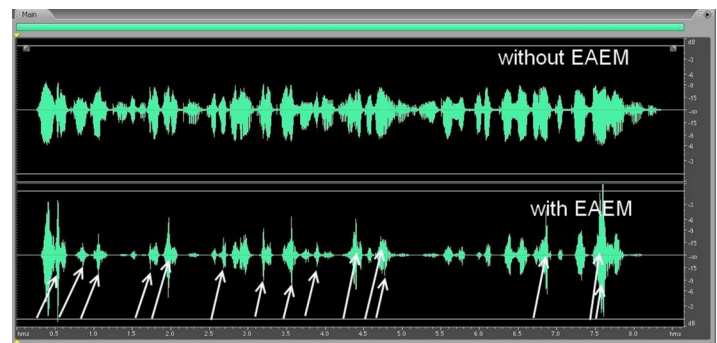


Figure 4. The TV commercial example. The upper waveform is the manipulated original, and the lower waveform is after the first step EAEM application.

Fig. 4 is the TV commercial example: Recording from the source commercial is digital. Inter-word and inter-sentence audio gaps are cut and removed. Inter-word gaps are very short or totally removed. Such manipulations may also be detected linguistically, but in that case the application of EAEM in parallel provides a scientific indication and may identify the exact instances of the manipulations. The upper waveform is the manipulated original, and the lower waveform is after the first step EAEM application. No clear marks are visible before the EAEM application but some have emerged after EAEM. Suspect locations are marked with arrow markers on the EAEM applied waveform: 9 of the 15 marked suspects are actual manipulations, the remaining 6 (false positives) are due to particular natural audio properties. Second step EAEM includes expanding and hard limiting, so waveform view will not be useful, so we have to continue with the spectrographic view.

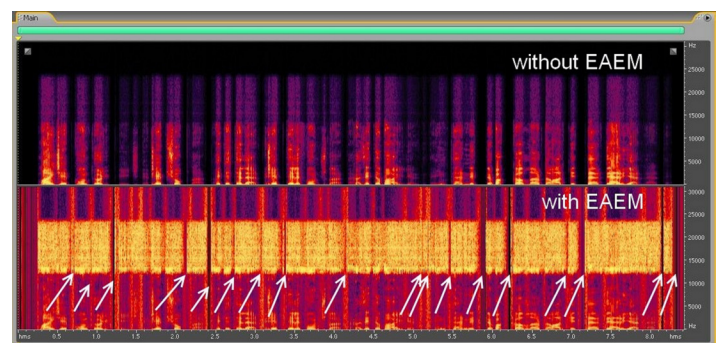


Figure 5. This figure shows the 2nd step of the EAEM applied to the audio of Fig.4. Top spectrogram is the expanded original, and the lower one is after applying EAEM step 2

In Fig. 5, we see the 2nd step of the EAEM applied to the audio of Fig.4. The 2nd step enables us to clearly identify suspect points (marked with arrows) including those missed at the first step. We see 18 suspected zones. The actual number of manipulations is 20. The 2 misses are later identified in lengthy waveform analysis at 0.5 and 1.7s.

Discussion

Most of the time, when any audio editing is made, one or more audio files are combined, cut or manipulated in some way, and original files may not be accessible. Such audio files that may be used as evidence are not accepted by the courts because of the editing at one or more locations. However, in crimes committed against the state in particular, the edited audio may have authentic sections where the speech content has not changed and can be

used, although the recording considered as a whole is no longer authentic. For this reason, it is necessary to identify both the edited segments and the unedited segments where the speech content has not changed.

Most of the time, uncertainties can occur in the decision-making processes that require long-term detailed examinations and can cause mental fatigue of the analyst. Mental fatigue can cause forensic audio analysts to make vital erroneous decisions.

Thus, the Enhanced Audio files Examination Method (EAEM) developed in the Forensic Audio Laboratory of Istanbul University Forensic Sciences Institute is important in aiding the analyst, and is being used in the same institution as a supporting method, since 1998 [4].

The manipulations of digitized sound files, which are the subjects of Forensic Audio Examinations, are done for a purpose. Digitized audio files are edited for morally reprehensible purposes such as slandering other people or institutions by changing the locations of sentences and words, thus causing an important difference in the content of the conversations, and misleading the listeners. Furthermore, words may be cut into smaller sound segments and then combined back together in a different order in order to create new words which will change the meaning of the conversation. Audio files generally contain acoustic background sounds, echo, reverberations, continuous electronic noise and important foreground talking sounds.

In such a case, the possibility of finding suitable zero-crossing or low energy points to cut is quite low for editing and alteration. In order to avoid detectible editing remnants, multiple cuts are applied to loud regions to minimize them. However, as a result of this process, the naturalness of the acoustical environment sounds and conversations are lost, and phonetically recognizable alteration remains and explicit differences in aural listening occur. When editing by adding and or subtracting sections is limited to only the optimum zero-crossings, editing deviates from its intended purpose and may not achieve the intended unlawful goal.

Conclusion

The standard conventional methods are: critical listening, waveform and spectrogram analysis, phonetic analysis. EAEM

does not replace, or remove the need for the conventional methods. Although the use of EAEM by forensic audio examiners seems to overload as an extra analysis, the extra parameters gained can be a significant time-saving factor in the decision phase.

While the developed technique is very useful, in practice some manipulations may be smoothened very successfully by the editing software, segments merged and fused with an insignificant mark and/or lost in the background noise. This reduces the success rate, and some marks that should appear can no longer be observed.

We used the FFT spectrogram in our visual frequency domain studies. However, our experience shows that the resulting FFT spectrogram cannot show all the sounds we can hear, for example, a whisper or breath. In some cases it appears that the signal is spread in the FFT. A sharp and distinct sound in the time domain is spread out in the spectrogram, and vice versa. A FFT spectrogram does not show time localization. But since we need speed in our studies, we generally use the FFT spectrogram. With higher computing power in the future, new signal processing algorithms may replace the FFT spectrogram and be able to remove this problem.

References

1. Steven W. Smith, The scientist and engineer's guide to digital signal processing, 2nd Ed, California Technical Publishing, USA, 1997. pp.141–155, 169–195, 218–231. Also available at: <http://www.dspguide.com/pdfbook.htm> accessed date 14.10.2017
2. Oppenheim A, Willsky AS. Signals and systems, Prentice Hall, 1983. New York, USA, 184-5.
3. Kamsi MA. 1998. Gibbs Phenomenon. <http://www.sosmath.com/fourier/fourier3/gibbs.html> accessed date n October 14.10.2017
4. Güral N. E. 2003. Bilgisayar Ortamında Kurgulanıp, Manyetik Banda Kaydedilen ses sinyallerinin adli amaçlı incelenme yöntemi, PhD Thesis, Istanbul University Institute of Forensic Sciences, Istanbul, 40, 90, 91
5. Adobe Audition 3 User Guide, Adobe Systems Incorporated. Also available at: http://help.adobe.com/archive/en_US/audition/3/audition_3_help.pdf accessed date 14.10.2017
6. Zakariah M, Khan MK, Malik H. Digital multimedia audio forensics: past, present and future, Multimed Tools Appl, <https://doi.org/10.1007/s11042-016-4277-2>, Springer US, New York 2017:2-4. accessed date 14.10.2017